A COMPARISON OF ACTIVATION FUNCTIONS FOR MULTILAYER PERCEPTRONS

By

SUDHEER NALLAVELLI

Bachelor of Technology

Regional Engineering College

Warangal, India

1997

Submitted to the Faculty of
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2001

A COMPARISON OF ACTIVATION

FUNCTIONS FOR MULTILAYER

PERCEPTRONS

Thesis Approved:

J. Chandler

Thesis Adviser

R. E. Hedrick

H. Lu

Timothy J. Pettibone

Dean of the Graduate College

PREFACE

This paper investigates the use of a shifted and scaled arctangent function as the activation function in a multilayer perceptron. It compares this activation function to the commonly used sigmoid function (Bishop, 1994) and the less common saturating linear function (Hagan, Demuth, & Beale, 1996). All optimizations are carried out using Richard Brent's derivative-free PRAXIS algorithm (Brent, 1973), a very robust procedure. Satisfactory results are obtained. It has been shown that the proposed activation function has equally good performance in convergence and attainment of global minimum to that of usual choice, the sigmoid function. The overall performance of arctangent function is better than sigmoid function.

I would like to express my heartfelt thanks to my thesis advisor, Dr. J. P. Chandler, for his guidance, expert advice in solving problems that I have come across in my thesis work. I also would like to thank Dr. G. E. Hedrick and Dr. H. K. Dai of the Department of Computer Science for serving as my thesis committee members. I very much appreciate their patience in reviewing and revising the report.

Finally, I would like to thank my parents, Sujatha Nallavelli and Srinivas Rao Nallavelli, for their sacrifices throughout their life to give me a better future. I deeply appreciate both my brothers, Ramana Nallavelli and Ravi Nallavelli, who extended their moral and financial support in gaining my advanced degree. I would like to thank my friends at

Oklahoma State University and in Dallas, particularly murali, jayanth, reddy, and sunil for helping me in achieving this goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

NOMENCLATURE

arctan	scaled arctangent function
ANN	artificial neural networks
FPRAX	final RMS error of the converged network
max(x,y)	function that returns the maximum of two input values x and y
min(x,y)	function that returns the minimum of two input values x and y
Ncalls	number of iterations made for convergence
PRAXIS	principal axis algorithm
RMS	root mean squared value
sig(x)	The sigmoid function
satlin(x)	The saturating linear function
WATFIV	Waterloo Fortran-IV compiler

1. INTRODUCTION

1.1 History

Neural networks are a class of machines influenced by Biology and Psychology. In the early 20^{th} century some work on learning, without any mathematical models, was done (Hagan *et al.*, 1996). In the 1940s Warren McCulloch and Walter Pitts demonstrated that artificial neural networks could compute any arithmetic function (1943). They modeled the human brain as a collection of interlinking threshold devices. These threshold devices are the counterparts of biological neurons and are called artificial neurons. Donald Hebb's learning mechanism followed this (1949). He recognized the conditioning properties in every biological neuron. Knowledge is stored in inter-neuron connection strengths known as synaptic weights. Frank Rosenblatt invented a perceptron learning network in the 1950s (1958). It was a success in the field of neural networks in certain respects: but there were shortcomings to these models. The paper published by Minsky and Papert (1969) criticized the developed models and questioned the future of neural networks. There was not much development in the field of neural networks during the 1970s. This is mainly because of the meager computing power and lack of new ideas, and perhaps the influence of Minsky and Papert. In the 1980s some new ideas reinvigorated the discipline. The backpropagation algorithm proposed by Rumelhart, Hinton, and Williams (1986) is one of them, although subsequently it has been pointed out that the method was actually contained in the earlier work of Werbos (1974). Since then artificial neural networks have been developed and used extensively. They have made their way into various commercial products and useful applications.

1.2 ANN Models and applications

Artificial neural networks are highly interconnected systems of neurons. These neurons are the processing elements in the neural networks. The fashion in which these interconnections are made will determine the network structure or model. The two most popular network models in use today are the feedforward network and feedback network models (Sarle, 1997). Other network models include Competitive and Dimension Reduction, to name two. New network models, as part of research, are found on a weekly basis. Some of these new networks are, sometimes, a result of minor modification to the well-known models. Depending on the problem at hand a particular network model is chosen. The multilayer perceptron, a fully connected network, is an example of a feedforward network. These have neurons in the form of a layer connecting to other neurons in the next layer in one single direction. Regression problems and classification problems (Sarle, 1997) use this type of network. In a feedback network, neuron connections between the layers form cycles. Output calculation is somewhat time-consuming because of the iterations introduced by the cycles. So is training. Bidirectional associative memory and Boltzmann machine are examples of feedback networks (Fausett, 1994). Neural networks, being complex in construction, have many applications in the real world. They are used in estimating the credit worthiness of an applicant. Aircraft simulators in the aerospace field, autograding of produce in agriculture, market analysis in securities, and voice recognition are some of the applications of neural networks (Hagan *et al.*, 1996). All networks need training before they are actually put to use. To train these networks, a lot of example test data are used. The following chapter explains about training, also called network learning.

1.3 Learning in Neural Networks

Learning is the process by which the neural network gains knowledge of the underlying patterns in the input data. Supervised learning and unsupervised learning are the two most commonly used methods (Sarle, 1997). In supervised learning, the input data and the expected output are both known. Input data is fed to the network and the output is compared with the expected result. The difference is used to adjust the network parameters in order to improve the performance. This process is repeated for the whole training dataset. Once the training is complete, test input data is fed in. The trained network will then calculate the output based on the experience/knowledge gained. Supervised learning is sometimes called "learning with a teacher". In unsupervised learning, the input data is known but not the target output. There are various types of learning methods that use unsupervised learning as the underlying scheme. Competitive learning (Fausett, 1994) and Kohonen networks (Kohonen, 1995) are a few examples. The most commonly used form of competitive learning is vector quantization (Kosko, 1992). In vector quantization the competitive units are related to a cluster center. The error function is formulated as the sum of the squares of the Euclidean distances between each training case and the nearest center. As this report concentrates on networks that use supervised learning, more information on unsupervised learning is not provided here.

Supervised learning is used for training the feedforward neural network. The following chapter discusses the structure of such network.

1.4 Feedforward Neural Networks

Feedforward networks are a valid generalization of their counterparts, the single-layer networks. Multilayer feedforward networks with at least one hidden layer can approximate any function to arbitrary accuracy (Hornik, Stinchcombe, & White, 1989). Any network with arbitrary connections can be represented by a mapping function. The main constraint on the network is that there should not be any feedback loops. It will help to represent the network as a function of connection weights and inputs. Given below is the figure of a feedforward network.

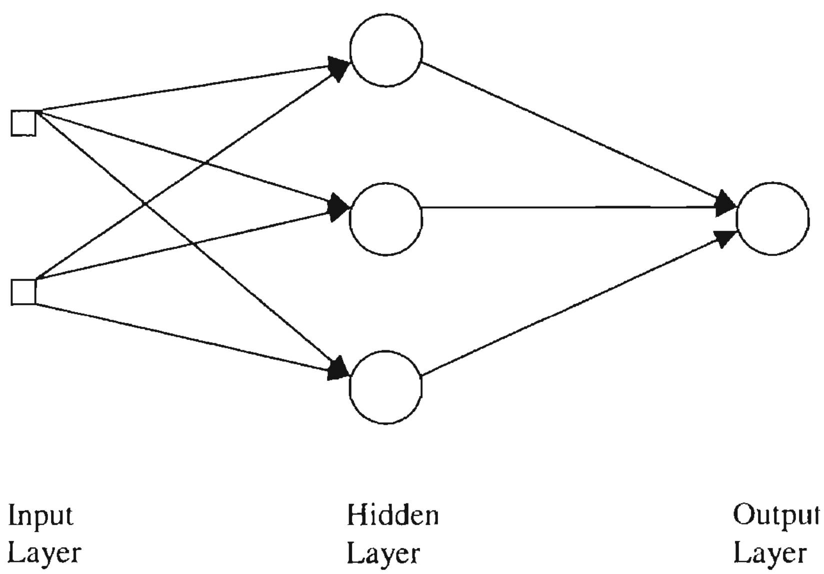


Figure 1.4.1: Multilayer Perceptron

The network shown above has three layers i.e. input layer, hidden layer and output layer. Each of these layers has neurons that are connected by weighted links to other neurons in the next layer. Following is the figure that shows a neuron.

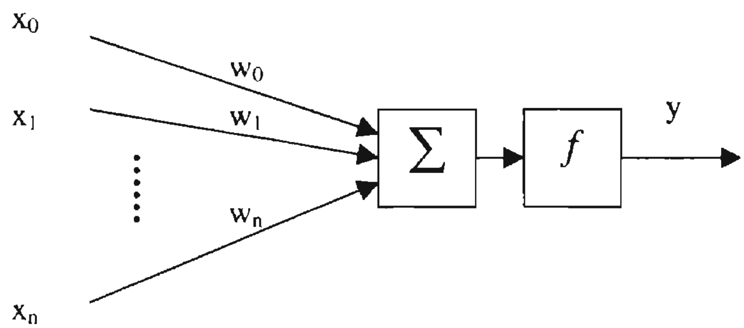


Figure 1.4.2 Single neuron model

The output of the neuron is given by a function of a weighted linear combination of inputs and the bias. Hence the output of the ith neuron is given by the following equation.

$$y = f\left(\sum_{i=1}^{n} w_i x_i\right)$$

Here f is a linear or nonlinear activation function. It may or may not be applied at the output depending on the type of problem dataset used. For a regression type dataset a linear function is usually used at the output layer (Bishop, 1994). These functions influence the network performance.

1.5 Role of Activation Functions in ANN

Activation functions introduce nonlinearity into the network. It is the nonlinear activation functions used in the hidden units that make the multilayer networks more powerful. A multilayer network with linear functions in the hidden nodes is similar to a single-layer network. This is because a linear function of a linear function is again a linear function. In a three-layered network any Tauber-Wiener activation function will fit any continuous output function (Chen & Chen, 1995). For a backpropagation training algorithm, however, the activation function needs to be differentiable. In practice a sigmoid function is used as activation function in the hidden nodes (Chen & Chen, 1995). It is differentiable and the output is bounded. Training with a sigmoid function is easier compared to a threshold function. The output of a sigmoid function varies with a small change in the input but a threshold function output will not change so easily. Depending on the targets, the function output can be scaled or vice versa. The output range of a standard sigmoid function is from 0 to 1. These bounded activation functions are sometimes called "squashing functions" (Hornik *et al.*, 1989). Activation function output can be modified to give negative and positive values. An activation function that gives both positive and negative values will learn faster when compared to the ones with only positive values because of better conditioning (Hagan *et al.*, 1996). Positive activation functions are used in this thesis because they are much more common.

1.6 Purpose

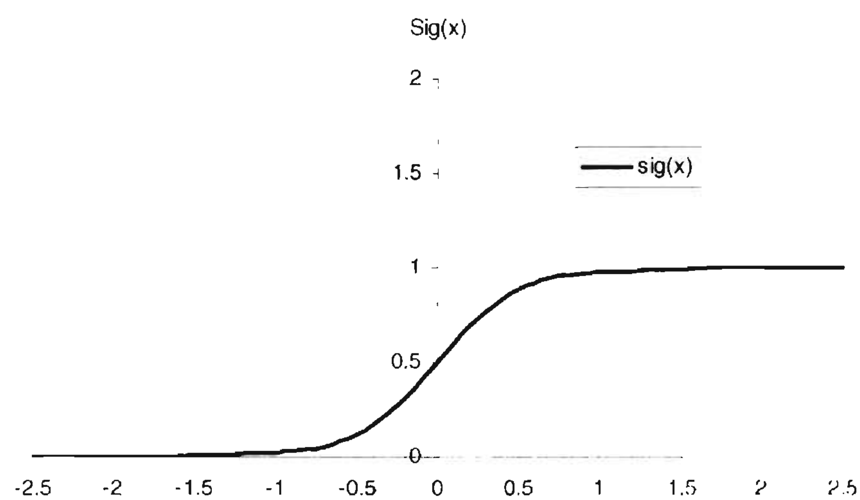
The main purpose of this report is to compare new activation function, or at least one rarely used, arctan(x), to the standard sigmoid and saturating linear functions. The motivation for comparing sigmoid with arctan is the function saturation point. Scaling of 'x' in each function has been chosen so that $f'(0)=1$. In backpropagation training, certain sigmoidal transfer functions may perform well with respect to computability and training times over others. This thesis will compare these three activation functions on various performance factors. The emphasis in this thesis will be on regression, fitting models to data by the least squares criterion, rather than on pattern classification, although one classification problem will also be tested. Several standard datasets from various repositories will be used for testing.

2. ACTIVATION FUNCTIONS

2.1 Sigmoid

The sigmoid function is also called as "log-sigmoid" transfer function (Hagan *et al.*, 1996). It is nothing but a linear transformation of tanh(2x). Hyperbolic sigmoids are a natural generalization of tanh(x) (Menon, Mehrotra, Mohan & Ranka, 1996). Figure below gives the specifications of this function (Abramowitz, 1965).

$$sig(x) = \frac{1}{1+e^{-4x}} = \frac{1+\tanh(2x)}{2}$$



X	Sig(x)
-3	6.14E-06
-2	0.000335
-1	0.017986
-0.9	0.026597
-0.8	0.039166
-0.7	0.057324
-0.6	0.083173
-0.5	0.119203
-0.4	0.167982
-0.3	0.231475
-0.2	0.310026
-0.1	0.401312
0	0.5
0.1	0.598688
0.2	0.689974
0.3	0.768525
0.4	0.832018
0.5	0.880797
0.6	0.916827
0.7	0.942676
0.8	0.960834
0.9	0.973403
1	0.982014
2	0.999665
3	0.999994

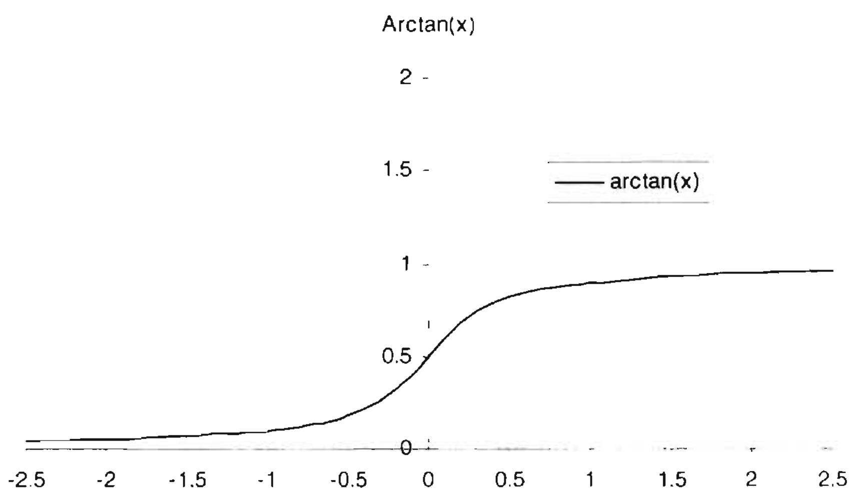
Figure 2.1.1 Sigmoid function graph

The equation given is a modified version that has a slope of unity at the origin. A network trained using tanh(x) as activation and one trained using sig(x) are the same; the difference is absorbed in the weights and biases (Bishop, 1994).

2.2 Arctan

The arctan function is the activation function of interest in this paper. Minor changes to the original function are done to maintain consistent slope at the origin. Equation and the graph are given in the following figure (Abramowitz, 1965).

$$F(x) = \frac{[a\tan(\pi * x) + \pi / 2]}{\pi}$$



X	F(x)
-3	0.033648
-2	0.050239
-1	0.098093
-0.9	0.108209
-0.8	0.120539
-0.7	0.135848
-0.6	0.155259
-0.5	0.180454
-0.4	0.213955
-0.3	0.259423
-0.2	0.321434
-0.1	0.403108
0	0.5
0.1	0.596892
0.2	0.678566
0.3	0.740577
0.4	0.786045
0.5	0.819546
0.6	0.844741
0.7	0.864152
0.8	0.879461
0.9	0.891791
1	0.901907
2	0.949761
3	0.966352

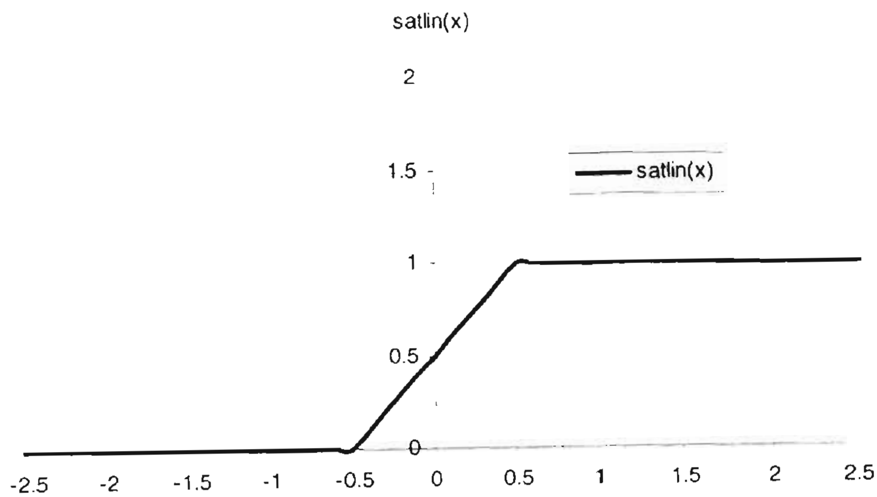
Figure 2.2.1 Arctan function graph

The arctan has the same characteristic 'S' shape as the sigmoid function.

2.3 Saturating Linear

This function has applications in hopfield networks (Hagan *et al.*, 1996). It is like a bridge between the threshold function and the sigmoid function. Observing the graph given below, this conclusion can be drawn (Abramowitz, 1965).

satlin(x) = max(0, min(1,x+0.5))



X	Satlin(x)
-3	0
-2	0
-1	0
-0.9	0
-0.8	0
-0.7	0
-0.6	0
-0.5	0
-0.4	0.1
-0.3	0.2
-0.2	0.3
-0.1	0.4
0	0.5
0.1	0.6
0.2	0.7
0.3	0.8
0.4	0.9
0.5	1
0.6	1
0.7	1
0.8	1
0.9	1
1	1
2	1
3	1

Figure 2.3.1 Saturating Linear function graph

2.4 Comparison graph

The graph given below is a combination of all the three functions listed so far. It will display the main motivation of this paper.

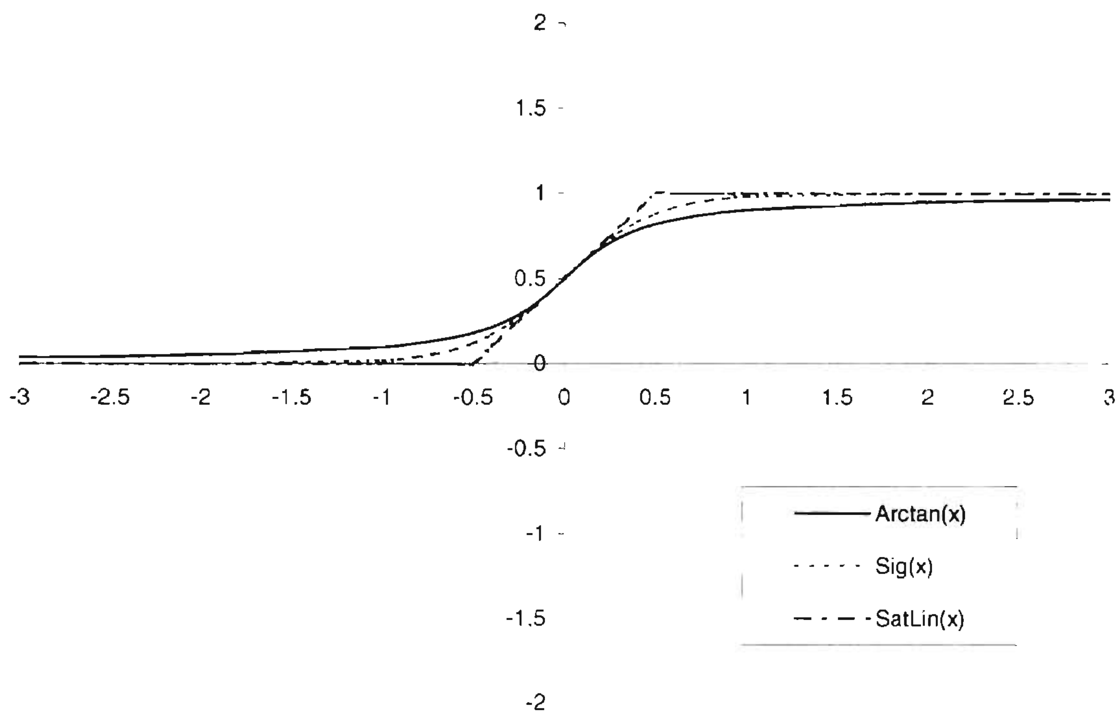


Figure 2.4.1 Comparison graph of three functions

A closer look at the function values between x=1 to x=2 on both positive and negative axes will give the saturation of sigmoid compared to arctan. Sig(x) output will be 1.0D0 in IEEE double precision in a typical computer when the input is ln(2^56)=38.8, whereas the output of arctan(x) doesn't saturate numerically until 2^56 i.e. 7.2D16. These calculations are based on a double precision floating point representation with 56 bits in the mantissa. The above difference in saturation point will influence the effect of large inputs on the sum of squares. This thesis will compare these three activation functions for speed of convergence, and the influence of outlier points on convergence. A check on convergence rate and the attainment of a global minimum will also be carried out.

3. COMPARISON METRICS

3.1 Convergence rate

Convergence is nothing but proceeding in a direction that reduces the mean squared error. But with a slow convergence rate the network has to undergo a lot of iterations before it reaches a minimum point. In the trajectory of this convergence there may be flat surfaces and valleys. In this paper, the factors influencing convergence rate are fixed. The influence of activation functions on the convergence rate will be tested.

3.2 Outlier points influence

The arctan activation function differs from the sigmoid in one aspect i.e. saturation. Figure 2.4.1 in Section 2.4 shows a comparison between these two. Data points that result in a larger weighted sum can contribute to the overall output. These data points, known as outlier points, should influence the performance of the network.

3.3 Global minimum attainment

The mean square error for an Adaline network (no hidden layer) has single minimum point (Hagan *et al.*, 1996). Hence convergence is guaranteed, provided the learning rate is small. In a multilayer perceptron there may be multiple local minima. So, global minimum attainment is not straightforward. To be certain about the convergence, a network should be trained at different initial conditions.

4. IMPLEMENTATION

4.1 Datasets

Carnegie Mellon University and University of Toronto's Data for Evaluating Learning in Valid Experiments (DELVE) maintains a repository of machine learning databases. This repository has a collection of neural network training datasets. Datasets "balloon", "bodyfat", and "wages" from the Carnegie Mellon University repository and "abalone", "bank", "boston", and "pumadyn" from the University of Toronto repository will be used in testing. The table below lists information on these datasets.

Table 1: Problem dataset specifications

Problem	Type of problem	Inputs	Outputs	Number of Examples
Abalone	Regression	8	1	800
Balloon	Regression	1	1	800
Bank	Regression	32	1	800
Bodyfat	Regression	14	1	250
Boston	Regression	13	1	506
Pumadyn	Regression	32	1	800
Wages	Regression	10	1	388

Every test problem has been treated as a regression problem, using a neural network with only a single output node and iterating until a local minimum was found in the sum of squares rather than stopping when all, or some specified number, of exemplars had been correctly classified, as is often done in classification problems (Bishop, 1994). For each of the above datasets, the network layout must be chosen. The "comp-activ" dataset was also considered, but was rejected on the grounds that it has missing data.

4.2 Regression vs. Classification

Classification problems have a predefined set of classes that the output should fall into. A network trained with this type of data will map the new inputs into one of the known classes. On the other hand regression problems try to fit or predict the values of continuous variables (Bishop, 1994). In a way regression and classification problems are special cases of function approximation. In regression problems the regression function will be approximated, whereas in classification problems the functions approximated are for the probability of membership in the output classes in terms of input variables. To test the influence of the activation functions in overall performance, regression datasets are chosen.

4.3 FORTRAN 77 to WATFIV

The base code is taken from Liya Wang's thesis "The damped Newton method – An ANN learning method". It has been modified to compile on the WATFIV compiler. Modifications to the program are listed below.

Positive array indexes

Network structure is stored in two two-dimensional arrays i.e. LAYER and WEIGHT. The row index is incremented by one in both these arrays. But, in the array LAYER, row –1 has been placed in row 13 and incrementing is done from row 0 onwards. Array LAYER stores the number of nodes in each layer and indexes to the nodes as well as their weights. The calculation of neuron and weight indexes has not been modified. Relevant loop variables are manipulated. If a bias has to be included in the calculation the loop variable will start from 1 instead of 2.

Similarly, array NEURON's row has been incremented by one. Row and column indexes of the three-dimensional arrays UUVV and HESIAN are incremented by one.

Learning algorithm

Wang's version of backpropagation has some incorrect second derivatives. Hence Brent's (1973) PRAXIS algorithm is used. It is much faster than backpropagation. The package minimizes continuously differential function using the principal axis method. It is derivative-free and robust. On Brent's test cases it was competitive with quasi-Newton methods, which are some of the best methods for unconstrained optimization.

Testing phase

In the testing phase the program reads the test inputs from an input file. It was reading the inputs to the network and not the relevant expected output.

Limitations of the program

The program will run with 800 training and test examples. It will store a 3-layer network (including input layer).

Zeroing out uninitialized values

The WATFIV complier does not allow accessing an uninitialized value. So, a few array locations that would never get initialized but are accessed are initialized to zero. Sections modified in this way are the bias locations in arrays UUVV, HESIAN and NEURON. All such modifications are commented appropriately.

- Bias in the output layer is simply a placeholder. It will not be initialized. Hence, a statement NEURON (all bias indices, 2) = 0 in a loop would set all the bias node first derivatives to zero.
- Function COMHAS accesses UUVV (*, *, 2) starting from the hidden layer for all nodes including bias. The array UUVV stores the partials of the weighted sum function of a neuron w.r.t the weighted sum of another neuron in its input path. Elements UUVV (all bias indices from hidden layer, all neurons including bias, 2) will be set to zero. Elements UUVV (all input layer nodes and bias, all neurons including bias, 2) are never initialized nor accessed.

4.4 Network

In order to decide the network layout the problem datasets are run on different layouts with varying numbers of nodes in hidden layer. Each time a node in the hidden layer is added, the weights from the previous fit with one fewer nodes are used. All seven problems are tested on network layouts with sigmoid activation function and with from one to five hidden nodes. A layout is finalized based on error improvement from the previous fit. The format of the finalized network layout is the number of nodes in input layer, hidden layer and output layer respectively. Table 2 given below lists the finalized network layout for each problem.

Table 2: Finalized network layouts

Problem	Finalized layout
Abalone	8-2-1
Balloon	1-5-1
Bank	32-2-1
Bodyfat	14-2-1
Boston	13-3-1
Pumadyn	32-1-1
Wages	10-3-1

4.5 Results

A finalized network layout for each dataset is trained using the three activation functions separately. Each network must be initialized with a set of initial connection weights. To initialize the connection weights a pseudorandom number generator is used. It will generate a unique series of pseudorandom numbers based on an input seed. Thereby the same set of pseudorandom numbers can be used as initial connection weights in separate computer runs. Maintaining a uniform training environment is an important step in comparing the activation functions. This section will list all the fifty fits, each from a different input seed, on every activation function for all the datasets. An asterisk (*) in the Ncalls column denotes non-convergence of the fit using that activation function. Column FPRAX gives the standard deviation (the RMS error) at the end of the fit. The RMS error of a nonconvergent fit is also listed for verification.

4.5.1 ABALONE dataset results

	Sigmoid		Arctan		Satlin	
Fit	Ncalls	FPRAX	Ncalls	FPRAX	Ncalls	FPRAX
1	191086	0.72326253269233387	34884	0.71736596036618749	7730	0.72514633289507391
2	265376	0.71569982912499475	27067	0.70293537026407293	10294	0.72514649515862750
3	500002*	0.70497788381144399	18159	0.71937383703663060	9965	0.72513995836268785
4	131472	0.71968542032505256	34371	0.70865991241031767	10001	0.72514059701500966
5	311670	0.72348633168516419	5804	0.71911253068829384	10310	0.72514626422824291
6	500065*	0.68851984684577394	11760	0.72204953236454272	963	0.77110443774839343
7	255155	0.72248488027283542	25387	0.71628043047569256	8684	0.72515170931243667
8	172997	0.72351894877376344	15688	0.71945031735722298	3954	0.73206282372741205
9	251948	0.72359904016631593	41182	0.71030352626197379	10694	0.72512802225719897
10	500027*	0.70228469766031854	27760	0.71566454303111138	6762	0.72513163403854464
11	500011*	0.71765544044809082	15799	0.71803319675278510	28583	0.71970517861874972
12	321971	0.70817667748644098	31171	0.71520948110905780	12727	0.72527395466735278
13	416502	0.72096360575234719	28861	0.70983076955203284	8529	0.72513990804535400
14	442340	0.71281710785244212	7424	0.71919783141667259	62592	0.71979189917379827
15	134519	0.72018549574448676	55996	0.69449642464811212	9610	0.72514037587184677
16	500014*	0.68442883840389312	8222	0.71979919683003868	1571	0.77110443774839277
17	500037*	0.71423658894284636	15136	0.71796592786777258	64878	0.70322200827072845
18	473577	0.72359847918389386	35261	0.69899257941732895	4492	0.73206282405252965
19	462383	0.70108938019906086	8877	0.72079923853727490	1471	0.77110443774839299
20	500012*	0.71246424391099139	12749	0.72498330093085095	59554	0.71975493436429538
21	299825	0.71786913095597737	8919	0.72203153457710156	37454	0.71971024060376942
22	265739	0.69364368815381505	19645	0.71723391030564343	1965	0.77110443774839255
23	500041*	0.71383508980340082	10291	0.72668169824681195	29409	0.72036062911914567
24	273043	0.71885707477797378	17513	0.72388537691849930	10084	0.72039212982772993

25	500039*	0.69435118598556722	17885	0.71941636847736623	1736	0.77110443774839343
26	500045*	0.69127615556079225	15897	0.71773121585471211	2232	0.77110443774839255
27	466857	0.71421058783414126	23719	0.71768582241270329	34589	0.71976649147661620
28	132718	0.70849513663714325	4234	0.72056722789434446	816	0.77110443774839377
29	500024*	0.69192097411684395	23515	0.71839881888241264	782	0.77110443774839277
30	308279	0.71424535982928883	9861	0.71942606476685045	45566	0.72063369189329285
31	217235	0.71243312952928262	24118	0.71943129032279507	6004	0.72524912246349260
32	500010*	0.69686035774211019	24519	0.71840657871090152	887	0.77110443774839410
33	439064	0.71165800950706670	16334	0.71793478764837648	20967	0.71983371086953973
34	500017*	0.70775715388267102	13739	0.72082330047523402	8516	0.72514683719361350
35	500049*	0.70232321210446902	14487	0.71922407893147033	785	0.77110443774839332
36	500018*	0.70229878903091403	10902	0.71820128373608960	9409	0.72514637865678511
37	343950	0.72018741208015014	11695	0.72511682205941341	10672	0.72514129043636655
38	500049*	0.68800716717762611	31111	0.71042487210298355	1300	0.77110443774839299
39	500015*	0.71042511943749298	5376	0.72086755200956387	9684	0.72513984356872685
40	190004	0.72018466398388847	13713	0.71974220535991784	7253	0.72514666626442503
41	262687	0.71763728741387756	37029	0.71475058147763137	6572	0.72514034275473371
42	96241	0.71863152709754841	39054	0.71005350651359567	7536	0.72513998560654402
43	416277	0.71673630402034672	28320	0.71708039712129601	16169	0.71971012368540244
44	500020*	0.70140535275834004	29311	0.71002556803204375	9758	0.72514629033743927
45	302111	0.71870773222506390	24457	0.71497390319339404	8067	0.72514678596782811
46	283340	0.72064050406723312	15046	0.70095120590256910	16940	0.72514929619128332
47	305704	0.72284638163777082	22949	0.70795174615729273	6725	0.72514040569028060
48	458926	0.72359831744026548	23763	0.71422376139559240	16735	0.72513997046247791
49	119152	0.71760220117538442	55714	0.69791153870145317	10153	0.72513989238522625
50	190399	0.71945691564345904	104628	0.70809706142308904	10504	0.72514011738731832

4.5.2 BALLOON dataset results

	Sigmoid		Arctan		Satlin	
Fit	Ncalls	FPRAX	Ncalls	FPRAX	Ncalls	FPRAX
1	29307	5.85960305946338286E-02	129482	5.63027797516419543E-02	439	0.13225328874402204
2	108235	5.62861365811446607E-02	9273	5.67979804666143701E-02	2153	6.01286139346198859E-02
3	5258	5.89825994709394855E-02	35619	5.64623630339680868E-02	1853	6.84439085709881295E-02
4	50520	5.67642786315073547E-02	16979	5.68848165502672670E-02	1360	6.01328954123945200E-02
5	196772	5.63380408570009344E-02	87223	5.74450678110472665E-02	2197	6.18845913687094754E-02
6	32954	5.79225428989072755E-02	17918	5.72887729692833358E-02	3634	5.90705446259457770E-02
7	16904	5.80705419021160241E-02	21640	5.63567483886834933E-02	2758	6.00603366410153994E-02
8	118628	5.74651112038544951E-02	11630	5.64251659995662763E-02	4624	5.84582352491938989E-02
9	114240	5.67501841841586557E-02	57362	5.74977203936818948E-02	1161	0.12415234976927476
10	22226	5.67721580703681727E-02	16519	5.65455191292083478E-02	7680	5.98073429449530458E-02
11	61440	5.83574862414223544E-02	36103	5.65115838321521291E-02	9926	5.81685762802519740E-02
12	90880	5.76222246905028282E-02	14078	5.63563011411535308E-02	3126	5.88786655897971384E-02
13	23527	5.63105014711305224E-02	18340	5.65216305977875519E-02	4429	5.78715064448045158E-02
14	73103	5.65996609581712953E-02	27883	5.66558145182137901E-02	2280	5.89493631761315168E-02
15	82036	5.65996214535263303E-02	25906	5.66262988543387927E-02	2034	6.18845913687093019E-02
16	132127	5.68219776487626227E-02	15109	5.64493027732429695E-02	2146	5.98719885774632518E-02
17	46928	5.74474182370280848E-02	12198	5.71507302691810076E-02	2520	5.90561716078503232E-02
18	61149	5.73716681509160928E-02	104501	5.61891843597623203E-02	6883	5.72360168556527646E-02
19	21057	5.88248032690392719E-02	18579	5.68046068563364120E-02	2537	6.17431513256344877E-02
20	35263	5.64042257333300656E-02	19314	5.79708860935993983E-02	1330	6.18845532306031981E-02
21	227017	5.65138162201765659E-02	22629	5.63151643863985044E-02	3396	5.84017133952426729E-02
22	62997	5.80500944674430877E-02	39940	5.62844287935416049E-02	2147	6.01328954123918000E-02
23	71700	5.64518214283199204E-02	152980	5.63382614673932289E-02	2682	6.18863875934285573E-02

24	17712	5.67649161537126593E-02	12901	5.69761405006310451E-02	2356	6.18845913687099611E-02
25	90959	5.63017044500956632E-02	31518	5.66278432222203920E-02	2201	6.18845913687095725E-02
26	154491	5.69642917949045244E-02	13390	5.65402765126379533E-02	1575	6.18845913687093505E-02
27	166049	5.73706253638637739E-02	35613	5.65343371584419510E-02	3576	6.01286139090986965E-02
28	58218	5.66453646373948072E-02	19882	5.66506390481397240E-02	3201	5.78334277464799471E-02
29	87973	5.76151045869480183E-02	23759	5.64184331714425585E-02	2893	5.89209531643439047E-02
30	39892	5.63102189228950056E-02	21357	5.64320891614770795E-02	2986	6.01328954123916751E-02
31	75080	5.62884282049249882E-02	64176	5.62819633112161613E-02	1325	6.18845913687093088E-02
32	98342	5.68357287839532446E-02	144320	5.76208896926978312E-02	5332	5.75549316514864934E-02
33	2668	5.89825994707197584E-02	18408	5.66121692756693176E-02	1779	6.17947320844742937E-02
34	49382	5.65996721803583847E-02	49141	5.73900912916055009E-02	1576	6.18845913687094892E-02
35	64783	5.62852729300987800E-02	7215	5.73315970295813770E-02	4590	5.91936912525612424E-02
36	94435	5.61667171323603151E-02	23199	5.65729708160880099E-02	3971	5.86325404242800177E-02
37	18676	5.84290652865113933E-02	19711	5.75880905898322054E-02	2482	6.01328954123918416E-02
38	81167	5.67079238740054170E-02	68310	5.68595990111095059E-02	1544	6.01125533839066348E-02
39	111095	5.74474097493700986E-02	40917	5.74369535076159968E-02	4162	5.89171942628364670E-02
40	123433	5.68066129222150676E-02	101132	5.67644794892833496E-02	1969	6.18845913687093574E-02
41	43218	5.62991360549696535E-02	18154	5.65275376088829884E-02	10372	5.74754720937490360E-02
42	38452	5.72921484744765702E-02	52437	5.58449535645137995E-02	2400	6.00959589269117536E-02
43	23703	5.76734569498561500E-02	161961	5.62013001003690282E-02	2222	6.18845913687093366E-02
44	37448	5.64042257333296979E-02	11684	5.66030608770637170E-02	1688	6.01328954125680826E-02
45	62124	5.66445892991151306E-02	110024	5.68119874529724317E-02	2036	6.18430631174693990E-02
46	15461	5.86821058302499921E-02	16992	5.65403784411383295E-02	2613	6.01328954123919526E-02
47	125171	5.62754893569098746E-02	22049	5.64889792581264524E-02	2132	6.18845913687096419E-02
48	68867	5.68075947481252799E-02	27265	5.78385979140008180E-02	2260	6.18427227434871496E-02
49	89988	5.60663793765045701E-02	45307	5.67171936353720388E-02	6068	5.97496185431042950E-02
50	68825	5.73716677535083680E-02	30753	5.75042452762339548E-02	4397	5.86833302493003917E-02

4.5.3 BANK dataset results

	Sigmoid		Arctan		Satlin	
Fit	Ncalls	FPRAX	Ncalls	FPRAX	Ncalls	FPRAX
1	500147*	0.10226549655743929	52086	9.76343128617629674E-02	3076	0.15047875170670258
2	44861	9.03827918550725445E-02	100991	9.53660283496702565E-02	33092	9.05983793540679899E-02
3	2541	0.15047875170670258	60157	0.10557091118425384	2596	0.15047875170670255
4	293419	0.10054682741525450	90954	9.33474306383126201E-02	46128	9.05965898776271011E-02
5	9180	0.15047875170670255	91749	9.11557816377930180E-02	4029	0.15047875170670258
6	500044*	9.73844823994552133E-02	64734	9.03117400684341071E-02	63936	9.05984115661955802E-02
7	3505	0.15047875170670263	207608	9.16532333000246308E-02	5775	0.15047875170670255
8	4213	0.15047875170670255	36831	0.10396737215542258	4253	0.15047875170670255
9	467084	0.10349108466368558	57749	9.42759530215667718E-02	70541	9.05965897368364304E-02
10	2788	0.15047875170670258	45459	0.10138141379508146	4508	0.15047875170670263
11	500003*	9.52728668248569027E-02	65045	9.03117400684340238E-02	203656	0.10483621384103585
12	5444	0.15047875170670252	114935	9.21817574749671520E-02	7250	0.15047875170670255
13	500077*	0.10373593815551811	58940	8.99834199911161148E-02	56555	9.05964389621789629E-02
14	448590	9.72882438447551373E-02	43050	9.25519386447355014E-02	42695	9.05983793540678095E-02
15	42483	9.03827918550725168E-02	48832	9.87156374395970204E-02	2805	0.15047875170670261
16	500030*	8.71857533747312091E-02	161902	8.73851534263826030E-02	183860	8.62198621454385861E-02
17	38785	9.03827918550726556E-02	122191	9.92202165345973941E-02	43039	9.06008811674696851E-02
18	500007*	8.76612973163178461E-02	102868	9.23596013336958482E-02	189875	8.47520592813001822E-02
19	500104*	9.53549658752815277E-02	75145	9.03117400684340516E-02	41889	9.05982936753728790E-02
20	41971	9.03827918550724613E-02	49379	9.03117400684342458E-02	39760	9.05982936753728929E-02
21	500014*	8.76962762401511664E-02	173863	9.00881434830272804E-02	153856	8.77308349559857426E-02
22	500019*	9.66718397543756408E-02	44954	9.03117400685450739E-02	48168	9.05965897368368606E-02
23	500070*	9.17532471916554793E-02	205989	8 96945235954331871E-02	136746	8.70676417966450428E-02
24	331154	0.10469823152250293	73161	9.17678014605497749E-02	41359	9.05983793540678789E-02
25	500105*	8.59730962963703577E-02	232518	9.16808294930443746E-02	163765	8.57127068604627146E-02

26	38576	9.03827918550725307E-02	54718	9.03117400684340932E-02	38377	9.05965897387917829E-02
27	40584	9.03827918550725307E-02	38723	9.03117400684340932E-02	39768	9.05964371747883246E-02
28	500127*	8.72565203425852559E-02	312846	9.00816291688477933E-02	205067	8.71535256075856996E-02
29	356114	0.10164271014859645	39444	9.69682886758148205E-02	45629	9.05983793540754007E-02
30	500049*	9.19649796341983627E-02	245648	8.65459356254367762E-02	120605	8.63190865578337130E-02
31	500127*	9.46420946138444297E-02	51262	9.03117400684466387E-02	65736	9.05984115467554918E-02
32	500028*	8.86374786806073123E-02	174662	9.08573659954104301E-02	164488	8.82908036502457888E-02
33	500000*	8.65194427603462185E-02	165686	9.18201680572649226E-02	130885	8.44011601779651038E-02
34	7664	0.15047875170670255	153632	8.60978670294414716E-02	4299	0.15047875170670255
35	43448	9.03827918550725584E-02	65611	9.29924271811897879E-02	61969	9.05984116870558187E-02
36	5239	0.15047875170670258	239710	9.03165370628914999E-02	2854	0.15047875170670261
37	420638	9.80188398898562407E-02	30403	9.30057562452517722E-02	40232	9.06008811674647585E-02
38	2518	0.15047875170670263	92558	0.10600471274733456	2834	0.15047875170670263
39	500116*	8.84079706155841694E-02	55365	9.38774555784888942E-02	64637	9.05984115542903812E-02
40	500118*	0.10092182061813737	81959	9.51748927646332288E-02	44948	9.05965899241672618E-02
41	451129	8.62834446707695935E-02	60077	0.10014593739023561	4012	0.15047875170670261
42	393637	0.10008590861796028	72841	9.03117400684341071E-02	57621	9.05964368222579863E-02
43	5218	0.15047875170670255	42234	0.10013373971143052	2581	0.15047875170670261
44	500110*	0.10374342832239093	46373	9.14833144161539408E-02	79557	9.71347768501526487E-02
45	4169	0.15047875170670255	60605	0.10465226824760977	7094	0.15047875170670255
46	3494	0.15047875170670261	97575	9.57134916995122514E-02	8238	0.15047875170670255
47	45432	9.03827918550855619E-02	50365	9.03117400684342042E-02	52026	9.05983793540677679E-02
48	6091	0.15047875170670255	52211	9 76397897370462203E-02	4286	0.15047875170670244
49	500096*	0.10157034573847354	65729	8.95544449455158709E-02	52250	9.05964373278495688E-02
50	3297	0.15047875170670263	165982	9.16925797065185344E-02	5501	0.15047875170670255

4.5.4 BODYFAT dataset results

	Sigmoid		Arctan		Satlin	
Fit	Ncalls	FPRAX	Ncalls	FPRAX	Ncalls	FPRAX
1	79953	1.64114278648292	276179	1.5478335199511373	16569	2.5967984319234572
2	7128	8.92946611252256	13826	6.7927293821335333	80480	1.3411853997850378
3	72984	1.32479176632722	101888	1.4335427068578115	11483	7.2232178983576629
4	93439	1.73453664312148	25783	5.9790821118510111	24794	6.1234543562528829
5	1869	8.94633245144607	4348	8.9463324514460716	4197	8.9463324514460734
6	168998	1.26238198864426	38759	5.8781445783296622	48513	3.0187330677641402
7	109663	1.63006376156130	57587	6.1673938536468000	28231	3.2463543509840838
8	1888	8.94633245144607	2687	8.9463324514460734	3519	8.9463324514460734
9	65256	1.43575312354761	24106	5.9782934340424285	83590	1.5138611154983099
10	96751	1.42611766150672	2715	8.9463324514460734	3097	8.9463324514460734
11	156618	1.61676103339509	43303	5.6907210924075704	17910	6.2637887059160393
12	96580	1.38560597431009	86271	2.5846352578792087	41235	2.4907705831024507
13	145395	1.53462865531371	2710	8.9463324514460734	4391	8.9463324514460734
14	101663	1.25786741735869	28892	2.1944920671937775	20096	7.0721054735009714
15	33959	1.82421457230264	4496	8.9463324514460698	55099	1.6717047086949597
16	91425	1.49147035954880	32249	5.3078298440949929	6612	8.3981136669906160
17	71942	1.40003828760744	238683	2.2521270054721483	31897	3.1727041082887006
18	145248	1.43815202439068	69060	1.6968949369822490	15509	6.3017817965373268
19	3481	8.94633245144607	4385	8.9463324514460734	3349	8.9463324514460716
20	46512	1.83242428862209	46152	5.9654196152332304	25828	6.2771659282607271
21	80510	1.59049307826431	60544	5.9398239324821533	24659	1.7261928162641713
22	53075	6.08224518188684	32010	6.0384648769132987	11685	4.1191099320742888
23	163379	1.37114097572671	18387	6.2096185237287145	10495	6.6375680805044359
24	1529	8.94633245144607	4600	8.9463324514460734	3556	8.9463324514460734
25	118506	1.38307071463466	3346	8.9463324514460734	3219	8.9463324514460734

26	148191	1.81055961139296	92828	2.3120260349479205	35385	2.7212354674992181
27	141504	5.91803939943730	87579	1.5259172923866362	34552	2.3450785219353865
28	84344	1.63387742714181	39239	6.2918908909032689	33560	1.6744511971204923
29	38490	6.32610253861794	71400	1.5354729612443172	22394	7.0384067147599243
30	192867	6.05200235476053	128149	1.6686924975413411	13396	7.8191825087212825
31	71702	1.41520240876895	128042	2.0388096683234749	21789	2.9830306615411284
32	189003	1.34844207438128	49515	4.4096320148373325	25916	2.7911241430328912
33	61276	5.36519569143911	2668	8.9463324514460734	4604	8.9463324514460734
34	90508	1.29395374454572	17920	6.4302529544020386	9495	7.0759508124285908
35	112644	1.51955227234221	137449	1.5957115478016550	17932	2.9497487521125203
36	43377	6.16764442067850	27550	6.0108672332352118	28242	5.2000970300466030
37	72522	1.66532465259441	136298	1.4747129348504273	17372	3.1873851292201527
38	110682	1.39818796567861	2786	8.9463324514460734	4158	8.9463324514460734
39	84575	6.12611662041725	3384	8.9463324514460734	4273	8.9463324514460734
40	125642	1.49900934543605	90059	1.6028253162205741	14393	3.8571616965630202
41	30414	6.48581196998823	3029	8.9463324514460716	44274	2.8433454642367622
42	51915	1.47303188857364	26794	6.0053423649358848	98187	1.2029383620738321
43	79428	1.49339996249187	34348	1.8104552113295800	74128	1.3917465094492725
44	91131	1.39883962052965	3085	8.9463324514460734	2929	8.9463324514460734
45	112270	1.56437000652485	154289	2.5860745423360294	34805	3.4315864030992334
46	148840	1.39816831318112	14379	5.7724345675925184	26566	3.6133669338100685
47	102662	1.36039402706386	116471	1.6002973576454693	6153	7.9437281102825565
48	78909	6.24976617901959	27490	5.8280933994064910	74678	1.5352170130591665
49	134801	5.68202325476004	2385	8.9463324514460734	15200	6.5182275041561564
50	76961	2.17075328717052	73509	1.3794317219342174	6459	7.4284322284942510

4.5.5 BOSTON dataset results

	Sigmoid		Arctan		Satlin	
Fit	Ncalls	FPRAX	Ncalls	FPRAX	Ncalls	FPRAX
1	321384	3.8246562170218210	38386	5.6385343183957595	37197	5.3828623510566853
2	272579	4.8480402310726030	242438	3.9159591829942433	32582	4.9968255346554722
3	271281	4.4156351295422080	39598	5.2678227839113481	5680	9.6311897909155118
4	110405	5.3421143681990637	61898	5.3510754983018227	58192	5.2853375962634397
5	194774	5.8857165571040460	67167	5.6983530381318976	21048	5.7142979472574229
6	136643	5.3414420630659656	62428	5.8794895266526677	26918	8.3599749503446361
7	331348	4.1849712195555915	93817	4.9988679303410013	56197	6.4453371080814756
8	500030**	4.7784422775231059	66489	5.0973124611266645	53909	5.7374817987065159
9	182682	4.4179644923244386	44260	7.6205133483756455	15570	8.3254855165307582
10	417000	4.0250686540595968	73259	6.4482578947996094	16788	6.9437847248081557
11	92032	4.2909063064192363	65661	6.2357791541417660	3708	9.6364678057772721
12	130637	5.3936785358831782	22107	8.4031580770085927	20370	8.5852970996126921
13	95435	5.9543997138977947	212391	4.1897660681764259	13070	9.1264196764866163
14	125241	5.3652682624437986	11886	9.2172504077107185	5659	9.6364678057772721
15	188268	4.3263062299617303	66255	6.2249462045193011	14901	8.6563070733034291
16	220271	5.3416731467890273	51844	8.1524452676640440	4045	9.6311897909155135
17	227986	6.4792413586865329	41182	6.5878315240098733	7000	9.6364678057772721
18	151468	3.9895133280956059	45533	7.8575081486147083	12152	8.6187539803386208
19	144012	4.2508177502533417	114977	4.2844278737884149	37216	6.1886232050783452
20	180389	4.2927048235220413	111459	5.4892791234591618	6748	9.6364678057772739
21	175964	4.3604028402467891	16769	8.6794549423165428	8982	9.2006648987188662
22	194580	4.3836186254104383	136462	4.4457139831588650	46994	7.7782822053694787
23	343693	4.1196820810326242	48821	9.0966570301951641	4592	9.6364678057772757
24	40409	5.3483385622507482	66663	6.4518650373287114	23044	8.5120441075497961
25	112333	5.0494689545120632	204309	5.8136093507287216	33402	6.8483086128731090

26	199329	4.7493069916467334	96261	5.9811963708370133	20508	9.1047262381980332
27	171347	4.0324222837990948	150301	5.4019217791936471	85292	6.0599418892600161
28	98636	5.2563582053863538	52027	5.2634526102606047	55104	5.5515218432943358
29	236205	3.5921202535055756	26704	6.4723453740961236	44385	5.9402333401318259
30	161319	3.9207449231102816	197285	4.0323329711331457	28469	6.6959942884816002
31	230805	4.4239168021592059	166480	4.1873063944133246	25303	6.4179268551871100
32	144550	4.2656901166121006	119117	5.0747214033352526	8659	7.8532177553595908
33	239305	3.8923646548952386	216286	4.2348738756392867	39084	5.9374443270394677
34	55219	5.6109775083483280	152129	4.9738660966543060	12891	9.1821614071092927
35	127504	4.5768159495237466	12626	9.2112145092650746	15657	8.4015720927844608
36	349199	3.8276397169970893	46111	5.3652558266282915	32130	6.5787139475867891
37	183216	4.3527214984993838	49978	8.3092124751307903	6608	9.6364678057772704
38	40248	5.9764037675004484	51795	8.4505605371468082	8395	9.2006648987188697
39	98650	4.4479312865093883	78034	4.6922065336123620	16540	7.3987153435871305
40	210830	5.6468679068935579	22073	8.0344777767791786	5860	9.6364678057772721
41	207369	5.7869883968786446	54788	7.1389335299348984	5868	9.6364678057772757
42	231309	4.1683884223034333	79854	4.9030200533252302	17542	8.1615373041440442
43	48845	7.0438781555878576	23519	6.0920535670469773	5814	9.6364678057772704
44	153502	4.5257291536611586	83309	7.3055920720315033	5915	9.6364678057772721
45	288686	3.7256518685571547	49731	6.0074267663860041	30219	6.5877986615800337
46	134086	4.3288214683033104	59837	5.5317730951579582	16556	8.1686798958579150
47	187209	3.9400387635656275	27738	7.6718011744774994	104275	5.5587822536218869
48	90349	5.2289942001978487	79368	5.2086247262228786	27773	5.1392915402092667
49	363826	5.3391812783632222	123571	4.9437384520692627	55768	6.0697221180905059
50	118966	5.5538506867129964	41370	6.2535236338133995	31270	6.0554677456341439

4.5.6 PUMADYN dataset results

	Sigmoid		Arctan		Satlin	
Fit	Ncalls	FPRAX	Ncalls	FPRAX	Ncalls	FPRAX
1	109182	2.08067089941857339E-02	43379	2.07597636701870140E-02	12092	2.16961436053599681E-02
2	151692	2.12477972843462710E-02	18009	2.06241109331604250E-02	43801	2.27090159043038808E-02
3	81884	2.12605594958669188E-02	67688	2.02155379483442560E-02	22423	2.13940133320403153E-02
4	117173	2.08355937311922937E-02	82496	2.12031793659528088E-02	44655	2.16542550170574326E-02
5	293406	2.16720446770588548E-02	61014	2.15482553972946890E-02	22903	2.19341426339849724E-02
6	263128	2.10754632600624950E-02	114793	2.02152627899908587E-02	8947	2.25240290530485283E-02
7	204019	2.07080870673381628E-02	58948	2.02156494452866223E-02	1562	2.25522453185695616E-02
8	103040	2.17100933620069936E-02	41269	2.20908178829757711E-02	2321	2.26197005233944974E-02
9	179992	2.13231741519668075E-02	27233	2.10442463985144705E-02	48815	2.14765779478278754E-02
10	72500	2.13651081888698649E-02	52326	2.14776966897201951E-02	10558	2.20894389299447214E-02
11	74963	2.17449448609070978E-02	31365	2.12301127607520262E-02	2939	2.25771185555180172E-02
12	112102	2.15604134494960659E-02	199717	2.02150764915349813E-02	22877	2.17300764205646500E-02
13	119440	2.20889947582724525E-02	40116	2.12189134844847638E-02	9943	2.24048714312525959E-02
14	94157	2.21804184133866274E-02	87691	2.18628287442729531E-02	1940	2.22771520097013002E-02
15	271957	2.10132998480267695E-02	39166	2.10247117618517915E-02	32445	2.20735112474182288E-02
16	87602	2.10459545321996198E-02	34118	2.16523587449684758E-02	63354	2.18176074675174407E-02
17	195592	2.18042780738826047E-02	32978	2.17642899024929615E-02	1425	2.25483441398617139E-02
18	153489	2.13662011101195502E-02	198394	2.02150848669351429E-02	42604	2.16157023372196139E-02
19	154110	2.10361732711357441E-02	40938	2.10415558624273587E-02	16622	2.17260625559865909E-02
20	92790	2.18191807935473638E-02	57650	2.11891763733827870E-02	5876	2.29483934523459880E-02
21	174318	2.13326339609367886E-02	79505	2.13341525897229670E-02	25892	2.26025026892703769E-02
22	105127	2.12840412246311263E-02	12739	2.09520282593736243E-02	2145	2.25989196577538257E-02
23	163682	2.19234772440680785E-02	47802	2.13605071843391871E-02	3582	2.45246619559717979E-02
24	308922	2.17651017641038481E-02	83584	2.06371240172412707E-02	10609	2.24725964992191565E-02
25	64868	2.20935014764404060E-02	50733	2.09012806303152700E-02	29513	2.08853428319603247E-02

26	82232	2.10429349924702198E-02	32221	2.10631356939349097E-02	1720	2.25743665787307320E-02
27	239430	2.13720115822417546E-02	21953	2.12447923812275792E-02	60397	2.03966542248548398E-02
28	111544	2.12964372981243114E-02	52268	2.10750581677609274E-02	50910	2.08660186346037307E-02
29	179684	2.09570676893960188E-02	31027	2.19427536564552363E-02	2852	2.18009345815702200E-02
30	62482	2.16053860062116951E-02	32560	2.11895025910798854E-02	1701	2.26472985444450513E-02
31	163392	2.12716872571341864E-02	177405	2.02150766124413045E-02	47497	2.03898332280778305E-02
32	147137	2.15229280393113599E-02	31221	2.11255043130978183E-02	5702	2.25787002524381539E-02
33	163619	2.18034448565969236E-02	94589	2.11085693204454278E-02	1753	2.30433492448086363E-02
34	46796	2.11472008670719920E-02	26191	2.13129973208168559E-02	47253	2.10207723301445645E-02
35	162915	2.12201572797894381E-02	123117	2.09556239665126404E-02	1750	2.22891211293765497E-02
36	82942	2.11656553289052730E-02	39346	2.10310106362794902E-02	13183	2.24130278698660007E-02
37	144862	2.11457121724203216E-02	48979	2.06472229991265140E-02	13278	2.20057212728146587E-02
38	215897	2.07999158257746106E-02	92419	2.11949720187504138E-02	1683	2.24932114342294362E-02
39	96750	2.21755095439979712E-02	35654	2.18238222061577052E-02	3555	2.42052296302881879E-02
40	116919	2.14800063029691839E-02	154652	2.02150788771893082E-02	6806	2.19536434835875434E-02
41	116670	2.12192184392612157E-02	103915	2.09224728638069957E-02	3748	2.21527288702756397E-02
42	103151	2.21364385572156673E-02	57584	2.06277305043303244E-02	11670	2.20668158864385129E-02
43	357334	2.06792869786902762E-02	204851	2.02150774998627845E-02	41701	2.12512625093848051E-02
44	131204	2.11364037166853411E-02	30423	2.07845611679417196E-02	1357	2.23028574722060058E-02
45	154621	2.17553876546291887E-02	20317	2.18006710869478142E-02	45213	2.20872868052376138E-02
46	80874	2.07229981459327028E-02	171942	2.02150854633802585E-02	22456	2.12480257794957078E-02
47	73252	2.15154707791458573E-02	30487	2.07191929398726389E-02	1931	2.20728856822299847E-02
48	263337	2.10564338674120131E-02	36539	2.11816273706409292E-02	4682	2.24986363164230846E-02
49	385577	2.05936416930961631E-02	34380	2.06605988926416954E-02	56854	2.17651936815883643E-02
50	144856	2.12468348576012982E-02	45764	2.12027080211887667E-02	11518	2.22511086619893855E-02

4.5.7 WAGES dataset results

	Sigmoid		Arctan		Satlin	
Fit	Ncalls	FPRAX	Ncalls	FPRAX	Ncalls	FPRAX
1	500044*	3.9938775038694971	43320	4.2357745834640976	5233	4.8676378335842525
2	500010*	3.9797761874160189	107160	4.2818503517880790	16811	4.2677718634982904
3	153857	3.7701291192813167	42241	4.0877035844049932	34724	4.1153966207666670
4	500093*	3.9200213544815750	72104	4.2898899580208703	18632	4.8663158576861907
5	500015*	3.9893574746963907	25816	4.1955840642138087	33166	4.2445695291569994
6	246078	3.8342441898686368	23601	4.0502819045773304	13727	4.1944916461893014
7	500012*	3.9212067404362401	27011	4.2636754977559841	14346	4.8672946245211755
8	445957	3.8121138920665603	25506	4.2094172683961206	38568	4.4100375045393756
9	343162	3.9833642171119910	39364	4.2228026638837450	66460	4.0949544782108038
10	500028*	4.0015666960426763	38555	4.3421685353886028	12780	4.8672945159010235
11	500005*	3.7555503993772787	26149	4.1359168614423032	31595	4.1835667214297025
12	117817	3.7957518552003058	35830	4.2183740402842549	21578	4.2258021244842734
13	500002*	3.7042474000216421	68058	4.0540632394416951	25504	4.4888684081535075
14	500005*	3.9098390300408976	100424	4.1009535534070078	29664	4.2195735019241702
15	386239	3.7545395958704710	33856	4.3604887683871212	31441	4.2071590136733494
16	500032*	4.2225872580199999	28131	4.1931351319760113	3577	5.0674327690950394
17	317445	3.9730174061573398	52632	4.1422365656321940	14630	4.8671342160837163
18	500032*	4.0119823384311220	84319	4.1527118044718661	18165	4.3681857799339658
19	500089*	3.9755392432185239	39647	4.2889174421917247	25537	4.4458993304590768
20	275873	3.8812316824340467	89922	4.0587197701340019	31041	4.1558987351420322
21	500020*	3.8293474495062267	46970	4.1586533490323756	16805	4.2404826176256751
22	479972	3.9267336526208392	36706	4.4949164926696774	14940	4.8672956409438246
23	500088*	4.0025432887938797	46794	4.1063094332632790	35695	4.2464101199876829
24	318571	4.0236858130281341	59387	4.2031279193337054	23427	4.1830912167318894
25	307482	3.9853519192329387	23392	4.2564862544109383	3701	5.0674327690950411

26	271563	3.9596142423066354	58968	4.2875445121697284	2760	5.0674327690950438
27	334473	3.9854629667192403	44729	4.0101383432489861	15382	4.2518668485651325
28	500055*	4.0074593881762492	17513	4.3328499596673566	18649	4.5702235063632486
29	380793	4.2119420241526058	72108	4.1751883644829384	2805	5.0674327690950385
30	500014*	3.9164347574639788	71233	4.1454141974976793	16241	5.0324975875510729
31	241482	3.9804843427043313	39051	4.2260158855683736	2931	5.0674327690950411
32	500051*	3.9056286519941072	44979	3.9986420003364782	23379	4.3002403927101946
33	500070*	3.9733842023594841	38549	4.0665352798333334	19108	4.2102588052015868
34	206148	3.9983755157033234	45515	4.2602610552304796	5217	5.0674327690950394
35	500063*	3.9857388062026184	43320	4.2357745834640976	5233	4.8676378335842525
36	209481	4.0982799597430937	107160	4.2818503517880790	16811	4.2677718634982904
37	500028*	4.0171914036580239	42241	4.0877035844049932	34724	4.1153966207666670
38	169744	4.1660598181359862	72104	4.2898899580208703	18632	4.8663158576861907
39	189180	4.0542933237853607	25816	4.1955840642138087	33166	4.2445695291569994
40	500023*	3.9984049027371653	23601	4.0502819045773304	13727	4.1944916461893014
41	294126	3.8717206395749026	27011	4.2636754977559841	14346	4.8672946245211755
42	500033	3.8791870486758127	25506	4.2094172683961206	38568	4.4100375045393756
43	264150	3.8890729141291467	39364	4.2228026638837450	66460	4.0949544782108038
44	371270	4.0897369645777264	38555	4.3421685353886028	12780	4.8672945159010235
45	374163	4.0157611864345979	26149	4.1359168614423032	31595	4.1835667214297025
46	257769	3.9711251704920238	35830	4.2183740402842549	21578	4.2258021244842734
47	162712	4.1709143836992215	68058	4.0540632394416951	25504	4.4888684081535075
48	197027	4.1015896337320354	100424	4.1009535534070078	29664	4.2195735019241702
49	438553	3.9074220319876827	33856	4.3604887683871212	31441	4.2071590136733494
50	478680	4.0549894174626342	28131	4.1931351319760113	3577	5.0674327690950394

5. SUMMARY, CONCLUSIONS, AND FUTURE WORK

Training was carried out on fixed parameter basis. This is done so that performance of the trained network is solely based on characteristics of activation function. Section 4.5 lists the output for all the datasets. FPRAX column gives the RMS error at the end of the fit. Some of these fits have not converged even after 500000 iterations. Such fits are marked with asterisk (*) in Ncalls column. Final value of RMS error, converged or not, of sigmoid and arctan function are taken for each dataset and compared. These are captured under lower minimum value column. It will show the number of times each function has converged to a better ninimum than the other. Table 3 shows summary of comparisons.

Table 3: Performance summary

Dataset	Lower minimum value		Learning rate		Lowest minimum attainment
	sigmoid	arctan	sigmoid	arctan	
ABALONE	31	19	0	50	sigmoid
BALLOON	24	26	12	38	arctan
BANK	13	37	21	29	sigmoid
BODYFAT	36	14	14	36	sigmoid
BOSTON	42	8	6	44	sigmoid
PUMADYN	14	36	5	45	arctan
WAGES	46	4	0	50	sigmoid

Values in the Ncalls column are compared for learning rate. Smaller values signify faster convergence. In the above table, learning rate column provides this information. Under a comparison metric the values for a dataset add up to fifty, i.e. the number of fits performed. The better (best, lowest) minimum attainment is compared for each dataset. The smallest value of the RMS error on all fits for both activation functions is calculated. An interesting observation is the convergence of sigmoid function. For some of the initial conditions, it did not converge. The arctan function has converged in every instance on all the datasets. It is also inferred from the above table that the learning rate of the arctan function is better in all the cases. In terms of global minimum attainment the sigmoid function has better performance on five of the seven datasets.

Conclusion

The overall performance of the arctan function is satisfactory compared to the sigmoid function. It is recommended to use arctan as the activation function in problems that could incorporate influence of outlier points. The sigmoid function, which is better in most of the performance metrics, could be used when the global minimum attainment is necessary.

Future work

- The arctan activation function could be used with a different learning algorithm to analyze the outlier point influence.
- During training, the network using the sigmoid function showed slow convergence. The arctan function convergence was relatively fast. This could be used in an application where speedup in convergence is needed with no other parameter change.

- The arctan has a similar shape to that of the sigmoid function. The global minimum attainment of sigmoid is better than the arctan. The factors influencing this could be investigated.

REFERENCES

Abramowitz, M., and Stegun, I. A. (1965). *Handbook of Mathematical Functions*. Dover Publications, Inc., New York.

Anderson, J. A., and Rosenfeld, E., editors (1988). *Neurocomputing: Foundations of Research*, The MIT Press, Cambridge, Massachusetts.

Barron, A. R. (1993). "Universal Approximation Bounds for Superpositions of a Sigmoidal Function", *IEEE Transactions on Information Theory, Vol. 39, No. 3*, pp. 930-945, May, 1993.

Beale, R., and Jackson, T. (1990). *Neural Computing: An Introduction*, Adam Hilger, Philadelphia, Pennsylvania 19102.

Berndt, E. R. (1991). *The Practice of Econometrics*, New York: Addison-Wesley. http://lib.stat.cmu.edu/datasets/CPS_85_Wages

Bishop, C. M. (1994). *Neural Networks for Pattern Recognition*, Oxford University Press.

Brent, R.P. (1973). *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, New Jersey.

Chen, T., and Chen, H. (1995). "Universal Approximation on Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamical Systems", *IEEE Transactions on Neural Networks, Vol. 6, No.4,* pp. 911-917.

Choi, J. Y., and Choi, C. (1992). "Sensitivity Analysis of Multilayer Percepton with Differentiable Activation Functions", *IEEE Transactions on Neural Networks, Vol. 3, No.1,* pp. 101-107.

Davies, L. and Gather, U. (1993). "The Identification of Multiple Outliers", *StatLib-Datasets Archive.* http://lib.stat.cmu.edu/datasets/balloon

Fausett, L. (1994). *Fundamentals of Neural Networks.* Englewood Cliffs, New Jersey: Prentice Hall.

Freeman, J. A., and Skapura, D. M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques,* Addison-Wesley Publishing Company, New York.

Ghahramani, Z. (1996). "Pumadyn family of datasets", *Data for Evaluating Learning in Valid Experiments.* ftp://ftp.cs.toronto.edu/pub/neuron/delve/data/tarfiles/pumadyn-family

Hagan, M. T., Demuth, H. B., and Beale, M. (1996). *Neural Network Design*, PWS Publishing Company, Boston, Massachusetts.

Harrington, P. (1993). Sigmoid transfer functions in backpropagation neural networks. *Analyt. Chem.* 65(15), 2167-2168.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Upper Saddle River, New Jersey 07458.

Hebb, D. O. (1949). *The Organization of Behavior*, New York: Wiley.

Hornik, K., Stinchcombe, M., and White, H. (1989). "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks, Vol. 2,* pp. 359-366.

Johnson, R. W. (2001, June 1). "Lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men", *StatLib-Datasets Archive.* http://lib.stat.cmu.edu/datasets/bodyfat

Kasabov, N. K. (1996). *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering,* The MIT Press, Cambridge, Massachusetts.

Kohonen, T. (1995). *Self-Organizing Maps,* Berlin: Springer-Verlag.

Kosko, B. (1992). *Neural Networks and Fuzzy Systems,* Englewood Cliffs, New Jersey: Prentice-Hall.

Menon, A., Mehrotra, K., Mohan, C. K., and Ranka, S. (1996). "Characterization of a Class of Sigmoid Functions with Applications to Neural Networks", *Neural Networks, Vol. 9,* pp. 819-835.

Minsky, M. and Papert, S. (1969). *Perceptrons*, The MIT Press, Cambridge, Massachusetts.

Neal, R. M. (1996). "Housing in the Boston Massachusetts area", *Data for Evaluating Learning in Valid Experiments.* ftp://ftp.cs.toronto.edu/pub/neuron/delve/data/tarfiles/boston.tar.gz

Rasmussen, C. E. (1996). "Bank family of Datasets", *Data for Evaluating Learning in Valid Experiments.* ftp://ftp.cs.toronto.edu/pub/neuron/delve/data/tarfiles/bank-family

Revow, M. (1996). "Predict the age of abalone from physical measurements", *Data for Evaluating Learning in Valid Experiments.* ftp://ftp.cs.toronto.edu/pub/neuron/delve/data/tarfiles/abalone.tar.gz

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press.

Rosenblatt, F. (1958). "The perceptron: A probabilistic model for information storage and organization storage and organization in the brain", *Psychological Review, Vol. 65*, pp. 386-408.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). "Learning representations by back-propagating errors", *Nature* 323: pp. 533-536. Reprinted in Anderson & Rosenfeld (1988).

Sarle, W.S. (1997). Neural Network FAQ, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: ftp://ftp.sas.com/pub/neural/FAQ.html

Tollenaere, T. (1990). "SuperSAB: Fast Adaptive Back Propagation with Good Scaling Properties", *Neural Networks, Vol. 3,* pp. 561-573.

McCulloch, W., and Pitts, W. (1943). "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics, Vol. 5,* pp. 115-133.

Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences.* Ph.D. thesis, Harvard University, Boston, MA. Reprinted in Werbos (1994).

Werbos, P. J. (1994). *The Roots of Backpropagation. From Ordered Derivatives to Neural Networks and Political Forecasting.* New York: Wiley.

VITA

Sudheer Nallavelli

Candidate for the Degree of

Master of Science

Thesis: A COMPARISION OF ACTIVATION FUNCTIONS FOR MULTILAYER PERCEPTRONS

Major Field: Computer Science

Biographical:

Personal Data: Born in Warangal, India.

Education: Received Bachelor of Technology in Electronics and Telecommunications from Regional Engineering College in May 1997; Completed the requirements for the Master of Science degree in Computer Science at Oklahoma State University in December 2001.